DOCUMENT RESUME

ED 121 083 FL 007 546

AUTHOR Filipovic, Rudolf, Ed.
TITLE Studies 1. The Yugoslav Serbo-Croatian-English Contrastive Project.
INSTITUTION Center for Applied Linguistics, Washington, D.C.; Zagreb Univ. (Yugoslavia). Inst. of Linguistics.
PUB DATE 69
NOTE 39p.; For related documents, see ED 096 839, ED 108 465, ED 026 639 and FL 007 537-552; not available in hard copy due to marginal legibility of original document
AVAILABLE FROM Dorothy Rapp, Center for Applied Linguistics, 1611 N. Kent St., Arlington, Virginia, 22209 ($1.00)

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS *Contrastive Linguistics; Descriptive Linguistics; *English; Language Patterns; Language Research; Linguistic Competence; Linguistic Performance; *Serbo roatian; Slavic Languages; *Translation

ABSTRACT

The first volume in this series on Serbo-Croatian-English contrastive analysis contains four articles. They are: "Contrasting via Translation: Formal Correspondence vs. Translation Equivalence," by Vladimir Ivir; "Approach to Contrastive Analysis," by Leonardo Spalatin; and "The Choice of the Corpus for the Contrastive Analysis of Serbo-Croatian and English," by Rudolf Filipovic. The fourth article, "Approximative Systems of Foreign Language Learners," by William Nemser, is ED 026 639. (CLK)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED121083

THE YUGOSLAV SERBO-CROATIAN - ENGLISH
CONTRASTIVE PROJECT

Director: Professor Rudolf Filipović, Ph. D.

B. STUDIES

1

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

Center for Applied Linguistics

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

FL007546

ZAGREB, 1969.
INSTITUTE OF LINGUISTICS
Faculty of Philosophy, University of Zagreb
Zagreb, Yugoslavia
CENTER FOR APPLIED LINGUISTICS
Washington, D. C., USA

THE YUGOSLAV SERBO-CROATIAN - ENGLISH CONTRASTIVE PROJECT

B. Studies

I

William Nemser: Approximative Systems of Foreign Language Learners

Vladimir Ivir: Contrasting via Translation: Formal Correspondence vs. Translation Equivalence

Leonardo Spalatin: Approach to Contrastive Analysis

Rudolf Filipović: The Choice of the Corpus for the Contrastive Analysis of Serbo-Croatian and English

ED026 639

Edited by
Rudolf Filipović

Zagreb 1969

INSTITUTE OF LINGUISTICS
Faculty of Philosophy, University of Zagreb
Zagreb, Yugoslavia

CENTER FOR APPLIED LINGUISTICS
Washington, D.C.
USA

The Yugoslav Serbo-Croatian - English Contrastive Project is an international venture involving the cooperation of Yugoslav and American scholars. Its primary aim is to facilitate the teaching of English as the second language to speakers of Serbo-Croatian. The results should also have relevance for the teaching of Serbo-Croatian to English speakers. It is further hoped that these results will afford new insight into the linguistic structures of the two languages and will constitute a contribution to contrastive linguistics.

The Project is directed by Rudolf Filipović, Professor of English and Director of the Institute of Linguistics of the Faculty of Philosophy, University of Zagreb, Yugoslavia, and coordinated by the Center for Applied Linguistics, Washington, D.C., U.S.A., represented by William Nemser, Director of the Center's Foreign Language Program.

The Project is supported jointly by the governments of Yugoslavia and the United States, and by the Ford Foundation.

The results of the Project research are presented in three series: A. Reports; B. Studies; C. Pedagogical Materials.

Vladimir Ivir (Zagreb)

CONTRASTING VIA TRANSLATION: FORMAL CORRESPONDENCE VS. TRANSLATION EQUIVALENCE

0. Contrasting means bringing two language systems together, setting them against each other and, hopefully, drawing illuminating conclusions about their nature, mode of functioning and use. Translation, ideally, means bringing two texts into a relation of equivalence, so that the text in the target language carries an amount of information identical to that carried by the text in the source language.

The important point to note in those two definitions is that in the case of contrasting we are dealing with langue and in the case of translation with parole. This is the distinction that should always be present in the minds of those who choose to base their contrastive analysis on a corpus of material from one language and its translation in another.

0.1. There are several ways in which the contrastive analysis of two languages can be made in practice. One approach would consist in taking the grammatical descriptions of the two languages and contrasting them chapter by chapter, paragraph by paragraph, noting similarities and dissimilarities. This procedure sounds very appealing and is certainly the most satisfying provided several preconditions are fulfilled: first, it presupposes the existence of the two descriptions in written or any other form available to the analyst; second, these two descriptions would necessarily have to be based on the same model of language; and third, they would have to be equally exhaustive. One sees that the requirements for the practical implementation of this procedure are indeed formidable - so formidable, in fact, that they have never yet been fully met, nor are there any real chances of their being met in the near future.[1]

0.2. Recognizing their inability to satisfy such exacting requirements, most analysts have settled for what they regard as the second best method - that of relying on their intuition supported by a larger or smaller body of "examples" in the form of tables and paradigms from existing grammar-books or in the form of translated sentences designed to illustrate certain statements. It is noteworthy that the analyst's intuition and his examples

are very intimately connected with translation. The next logical step has been to aim at methodological rigour and exhaustiveness by eliminating intuition (with its chance results) and by working exclusively on a large corpus of material in the original language, its translation into the language with which it is contrasted, and its "back-translation" into the original.[2] This means that translation has not only been brought directly into contrastive analysis but that it has also been made into one of its cornerstones. It is precisely for this reason that a very real need is felt for a critical examination of the value of the linguistic material obtained by translation for the purposes of contrastive analysis.

1. We have already noted that translation deals with the physical realization of language, with parole, while what we contrast in contrastive analysis are language systems in abstract, i.e. langue. It is therefore our job as analysts, when using translated materials for contrastive analysis, to abstract the system from the text. In this case, we are in a position not unlike that of the grammarian trying to extract a valid description of a language from a large corpus of recorded linguistic material (hoping that his descriptive statement will also have certain generative properties). And we face the same limitations that he faces. No matter how big our corpus, we can never be sure that we have covered all relevant aspects, that we have not missed certain vital and very revealing points. But we also share with him the same advantages: the scope that we cover in this manner can be very broad and our conclusions are carefully documented with objective and reliable linguistic evidence; nor should one minimize the advantages in speed and accuracy that the possible mechanical processing of linguistic data offers in such a case.[3]

1. 1. Without at all questioning the feasibility of a project of contrastive analysis based on translated material,[4] let us now examine how and to what extent translation can serve as a starting point in contrasting two languages. But before we do this, an observation of a purely theoretical nature should be made: the very idea of contrastive analysis presupposes the existence of language universals of some kind (or at least of some elements common to the two languages under investigation). This holds good equally well when contrasting proceeds via translation. The analyst isolates certain elements of structure in one language (such as tenses, plural markers, possessives, demonstratives, word order, etc.) and observes what elements of structure correspond to them in the other under conditions of semantic equivalence.

An important difference between translation and contrastive analysis is that the former aims at establishing semantic equivalences between the two texts (without necessarily relying on consistent correspondences between the elements of structure involved) and the latter aims at establishing formal-semantic correspondences between the two systems (utilizing only those products of translation work which display clear and consistent formal features of correspondence).

Although what has been said here may seem rather far removed from any practical considerations, it is nevertheless very relevant for actual contrastive work because it boils down to the basic question of the units of contrastive analysis and the units of translation. It can be said that the units of (syntactic) contrastive analysis are the units of grammatical description in terms of which the two languages are presented. The units of translation are the elements of semantic content which can somehow be extracted from the linguistic material of the text in the source language and expressed in the target language using the linguistic tools of that language. (The implication here is that content can be separated from form.)

When one attempts to isolate the formal carriers of these elements of semantic content, one comes out with morphemes, words, groups, clauses, and sentences.[5] To these one should perhaps also add the paragraph and the whole work, which are sometimes regarded as the appropriate translation units.

1.2. Contrastive analysis would proceed most easily with translation equivalences if all translation was rank-bound at word or even morpheme rank. However, since normal translation - and that is the kind used in corpus-based contrastive analysis - is usually not rank-bound (especially not at morpheme or word rank), but is rather characterized by frequent rank shifting and changes of the ranks at which translation equivalences are found, the analyst faces considerable difficulties in trying to decide upon the elements of formal correspondence that can be usefully contrasted. The situation is further complicated by the fact, noted by J. Darbelnet[6], that paraphrase (including additions, deletions, restatements, etc.) is often involved, and that it is in the area between literalness and paraphrase that some of the best translations are to be found.

1.3. The question that this paper poses, and attempts to answer, is the

following: How much of the translated material produced by normal (unrestricted) translation can the contrastive analyst use? The question can be asked differently: How does he know which products of translation to eliminate as unsuitable for contrastive analysis? Or finally: How does he find formal correspondences between an original text and its translation? We shall assume that he works with a corpus and its translation and that he already has the units of the language of the original for which formal correspondences are now sought - the semantic equivalence having already been established - in the language of the translation. A formal correspondent has been definded by Catford as "any TL category (unit, class, structure, element of structure, etc.) which can be said to occupy, as nearly as possible, the 'same' place in the 'economy' of the TL as the given SL category occupies in the SL".[7]

2. It is important to realize that there are, in every text, a number of points at which formal grammatical correspondence can be established.

> There can be no doubt that the personality and the equilibrium of the individual are gravely threatened by technological civilization.
>
> Nema nikakve sumnje da je ličnost i ravnoteža pojedinca ozbiljno ugrožena tehničkom civilizacijom.

The corresponces represented here, excluding those lexical ones, include the following: E present tense - SC present, E negation - SC negation, E of-construction - SC genitive, E nouns - SC nouns, E singular noun - SC singular noun, E adverb - SC adverb, E passive - SC passive, E attributive adjective - SC attributive adjective, E by-agent - SC instrumental agent, E that-clause - SC da-clause, E number concord - SC absence of number concord, etc. As seen here, the correspondences are numerous and mostly in the same category. Yet, not all potential points of correspondence are utilized: the articles of the source sentence, for instance, seem to have no identifiable correspondents in the target sentence. A contrastive analyst, no matter what his immediate interest (except the article), would find this translation usable in his work.

But the source sentence could have a somewhat different target language rendering:

> Nema nikakve sumnje da tehnička civilizacija ozbiljno ugrožava ličnost i ravnotežu pojedinca.

This time the correspondence E passive - SC active is established, which is not without interest for the analyst contrasting the voice in English and Serbo-Croatian. At the same time, some other correspondences are changed: E subject - SC subject into E subject - SC object, E by-agent - SC instrumental agent into E by-agent - SC subject.

Other possible versions of the target sentence could be the following:

Nema nikakve sumnje da ličnosti i ravnoteži pojedinca ozbiljno prijeti tehnička civilizacija.

Nema nimalo sumnje da su ličnost i ravnoteža pojedinca ozbiljno ugrožene od strane tehničke civilizacije.

With each version, some of the correspondences are changed, even though the translation remains fairly literal (i.e. word-for-word). Changes become more drastic when the translation is free:

Sasvim je sigurno da tehnička civilizacija predstavlja ozbiljnu opasnost za ličnost i ravnotežu pojedinca.

Tehnička civilizacija nesumnjivo znači ozbiljnu prijetnju za ličnost i ravnotežu pojedinca.

Nesumnjivo je da tehnička civilizacija znači ozbiljnu prijetnju ličnosti i ravnoteži pojedinca.

But even at this stage, some correspondences are preserved, and the translation can still be used for certain aspects of contrastive analysis, e.g. E present tense - SC present, E adverb - SC adjective, E attributive adjective - SC attributive adjective, E that-clause - SC da-clause, E singular noun - SC singular noun, etc. However, for other areas of contrastive work, no usable correspondences can be found in the more free versions of this translation: thus, for instance, the analyst contrasting the negation, introductory *there*, the passive voice, agential expressions, or number concord in English with their Serbo-Croatian counterparts will be unable to use the last three renderings of the source sentence.

2.1. What all this points to is the fact that different translations of the same source text (all seemingly equally good equivalents) vary in the number and type of formal correspondences they enable one to establish and, consequently, in the kind of contrastive statement that can be made on the basis of these. (It can also be mentioned, in passing, that this fact would make the back-translation method unworkable: first, one would not know which of the several possible translations to "back-translate", and secondly, each of

these versions would produce several renderings, thus making the picture rather messy.)

3. It appears necessary, at this point, to expand Catford's definition of formal correspondence (see above) for purposes of contrastive analysis. Obviously, formal correspondence defined in terms of categories occupying identical positions (i.e. performing identical functions) in any two languages is both too rigid (such categories probably cannot be set up) and two narrow (it would prevent one from establishing many of the contrasts that the learner regularly establishes between his mother tongue and the foreign language). However, rather than rejecting the concept, one can modify it to include all those isolable elements (units, classes, structures, elements of structure, etc.) which can be said to occupy, as nearly as possible, the 'same' places in the 'economies' of the target text and the source text. Getting down to the "text", rather than language, enables one to establish correspondances at all those points at which there are, in the two languages, isolable formal carriers of semantic equivalence. This procedure is particularly relevant when contrastive analysis utilizes translated material, since it is the only one that makes allowances for the numerous cases of rank-and-level-shifting in translation. Thus, for instance, it was possible to establish the adjective--adverb correspondence in the examples in section 2 above, which Catford's definition would rule out.

3.1. A distinction should be made between formal correspondence and translation equivalence. While all cases of formal correspondence as conceived here are also translationally equivalent (naturally so, since they are based on translation), not all cases of translation equivalence[8] stand in a relation of formal correspondence; but formal correspondence being such a multi--layered phenomenon, it is difficult to find translation equivalents without a trace of correspondence - at least in languages as closely related as English and Serbo-Croatian. When the equivalent of "there can be no doubt" is taken to be "naravno je sigurno", then we can still establish the clause - clause and the present tense - present correspondences; even when the translation equivalent is reduced to "nesumnjivo", some correspondences remain (E clause - SC adverb, E negative adverb - SC negative prefix). And even in such extreme cases as idiomatic espressions and proverbs, which are of necessity translated quite freely, there are certain correspondences of which contrasts can be made:

Don't count your chickens before they are hatched.
Ne spremaj ražanj dok je zec u šumi. (E negative imperative - SC negative imperative)
It was raining cats and dogs.
Padalo (lijevalo, curilo) je kao iz kabla. (E preterite - SC perfekt, E continuous tense - SC imperfective aspect).

It follows from the nature of formal correspondence understood in this way, that not all correspondents should be sought in sentence-to-sentence relationships: translations that do not respect sentence boundaries but rather break sentences, fuse them or change their order will still yield many correspondences useful for contrastive analysis (except, of course, for the contrasting of the sentence structure itself):

All this leads to the realization that a "pedagogical grammar" cannot be the application of mere linguistic grammar, but is rather the result of combining linguistic grammar with psychological grammar.
Sve nas ovo dovodi do uvjerenja da "pedagoška gramatika" nije jednostavno primijenjena lingvistička gramatika. Nju ćemo dobiti tek kombiniranjem lingvističke gramatike sa psihološkom gramatikom.

(Although this translation of the second part of the original sentence fails to supply correspondences required for, say, sentence structure contrasting, it can still be used in contrasting adjective positions in English and Serbo-Croatian, and English gerund and Serbo-Croatian verbal noun usages.)

In the following example, too, the sentence structure is rather appreciably affected by the process of translation but there are nevertheless quite a number of correspondences because of which the example will be retained in our corpus and used for certain aspects of contrastive analysis of English and Serbo-Croatian:

It can be inferred from the foregoing that, if a change is to be carried through successfully, it must be acceptable to all people who will be affected by it, so it is the task of management to create the conditions that will ensure the greatest amount of support and the minimum amount of resistance.
Iz ovoga što smo dosad rekli proizlazi da se promjena može uspješno provesti samo onda ako je prihvatljiva svima koji će biti

njome zahvaćeni, pa je zato zadatak rukovodstva da stvori uvjete za maksimalnu podršku i minimalni otpor kolektiva.

3.2. This discussion has so far centered on the problems of translated corpus as a basis of contrastive work. It has shown that all translations, no matter how free, have their place in the corpus because formal correspondences can always be established between the source text and the target text. It has also shown that probably no translation can be made in such a way that correspondents of all isolable formal elements of the source text are found in the target text. Different translated versions of one and the same original text (all of them presumably equally valid as translation equivalents) will permit one to establish different formal correspondences.

4. All these considerations stem from what may be regarded as a macroscopic point of view of an entire contrastive project. When the focus of interest becomes more narrow and specific, however, when our perspective is that of an individual project worker dealing with a well defined problem rather than that of a project manager, the attitude to the corpus and its translation becomes different. First of all, the individual analyst will never handle the entire corpus but only that portion of it which contains the instances of his particular problem[9]. Second, he will only be interested in finding formal correspondents in the target text for the elements of structure of the source text which are the subject of his contribution to the project as a whole. And in trying to find these correspondents, he will see that a considerable proportion of translations simply fail to offer anything that he could accept as a formal correspondent of the element of structure that he is studying. (The exact proportion of such unacceptable examples will depend on how free the translation is and on the particular element of structure that is the subject of his interest. Some elements of structure seem to be more susceptible to "deviation" and perhaps disappearance in the process of translation than others.)

4.1. The problem that the analyst will have will be in the first place to determine whether a formal correspondent is present in the translated text or not - and if it is, what exactly it is. Experience will soon teach him that mere translation equivalence is not a valid basis for translating, and that extracting a formal signal of structure out of a text is in fact the first step that he should make. This will be easy in the language to which contrastive

analysis is directed and which is taken as a starting point. In this language, the elements of structure to be analyzed will be given by the description of that language. Let us assume that English is the language to which our contrastive analysis is directed, i.e. that we are trying to produce a contrastive grammar of English for Serbo-Croatian learners. In this case, a descriptive grammar of English will supply not only the actual topics for analysis but also their analytic breakdown and will give the analyst the metalanguage in terms of which he can discuss the contrastive features of the two languages. What this description will not give, however, is the list of the formal correspondents in Serbo-Croatian. These will have to be extracted from the semantic equivalences represented by the translated text, using the labels provided by the metalanguage adopted in the independent descriptions of English and Serbo-Croatian.

4.2. Suppose that the subject of contrastive analysis is the English participle. The grammar of English will state, and the corpus presumably confirm and exemplify, the fact that English has two participles, present and past, and that both of them display some adjectival and some verbal properties; it will also specify the conditions under which one or the other syntactic quality of the participle becomes prevalent. The analyst will then classify his participial examples into the two main groups and as many sub-groups as he feels his grammar can handle, or as many as are necessary for his purpose. Next, he will examine the translation equivalents in each particular sub-group, looking for the specific formal elements that can be said to correspond to the participles in the English original. In the sub-group containing -ing forms preceded by forms of the verb to be and followed by an object - where the syntactic interpretation of the -ing form is clearly verbal - he will find, for instance, Serbo-Croatian verbal tenses (of imperfective and, perhaps less obviously, perfective verbs). But he will also find Serbo-Croatian gerundives, as in the following example:

As he was making his way across the field, he ran into a...

Probijajući se poljima naletio je na...

Here he will begin to wonder whether the gerundive can be regarded as a formal correspondent of the English progressive preterite. But on closer inspection, the English example will reveal certain features which distinguish it from some other uses of the progressive preterite: first, it is used in a coordinate clause which opens the sentence; second, the clause could be reduced to a participle (Making his way across the field, he ran into a...);

third, the clause could be retained in translation, and the progressive preterite would have the Serbo-Croatian past tense of an imperfective verb as its formal correspondent (Dok se probijao poljima, naletio je na...). His conclusion may be that when all these conditions are fulfilled, the Serbo-Croatian gerundive is a proper correspondent (contrastively significant!) of the English progressive preterite. (As a matter of fact, the gerundive here stands as an equivalent and correspondent of the whole clause, not of its verbal part alone.)

However, some of the translations in the corpus will leave him without a formal correspondent, though obviously there will be formal elements carrying the meaning of the original:

> I paid McSwiggin and I paid him plenty, and I got what I was paying for.
> Gospodo, ja sam plaćao McSwiggina, i to šakom i kapom, a vjerujte da mi je to isplatilo.

It would be wrong to underline "se isplatilo" as a correspondent of "was paying" in this example. There is nothing in the translation that can be taken as a formal correspondent of the progressive tense. (The pastness is of course represented in the translation.) One could at this point legitimately ask whether the Serbo-Croatian text is a translation equivalent of the English text above; or in other words, whether the progressive tense has been rendered at all. One answer can be that the two texts are equivalent because its Serbo-Croatian part has been produced by a competent bilingual speaker of English and Serbo-Croatian. The second answer might be that they are not equivalent, that the meaning contributed by the progressive tense has not been included in the translation, and that it cannot normally be included in Serbo-Croatian.[10] Its inclusion, though possible, is not desirable, since it would affect the naturalness of the translation:

> Gospodo, ja sam plaćao McSwiggina šakom i kapom, ali sam i dobio ono (za) što sam plaćao.

Even when a formal correspondent seems to be present in the translation, care should be taken to see whether the correspondence is genuine or merely spurious, a result of mistranslation:

> Whenever he sees a picture of a big boat, he asks his mother if that boat is bringing Daddy home.
> Kad god negdje ugleda sliku kakvog velikog broda, zapita majku da li je to onaj kojim će se tata vratiti kući.

Finally, in some cases of unmistakable equivalence and clearly statable correspondence, it is difficult to see how these examples can be analyzed in contrastive terms:

> These booze customers had until then been buying their supplies from...
>
> Dotad su se oni opskrbljivali pićem kupujući ga od ...

Here the Serbo-Croatian present participle "kupujući" corresponds formally to "(had) been buying", but this fact is hardly relevant contrastively because it is an automatic consequence of the process of shifting in the rest of the sentence.

This example shows very well that formal correspondence is contrastively significant only when it is systematic (and perhaps also statistically frequent) and when it is achieved independently between the structures that are being contrasted rather than as a result of intra-sentence shifting which then affects also the correspondent structure in the translated text.

Another example of this situation is provided by the of-construction in the following sentence:

> One student could remember the order of all fifty-two freshly shuffled cards after his first twenty-minute study of them.
>
> Jedan student se mogao sjetiti poretka svih pedeset i dvije izmiješanih karata pošto ih je proučavao svega dvadeset minuta.

The accusative form of the personal pronoun ("ih") is the formal correspondent of the original of-construction ("of them"), but it is impossible to say whether this is an example of systematic correspondence or a chance result of the changes that the original sentence has undergone in the process of translation.

5. An attempt has been made in this paper to examine how translation can be used in contrastive analysis. It is claimed that contrasting via translation is possible but that the differences between the two should be carefully noted: the former is concerned with semantic equivalence, the latter with formal-semantic correspondence; the former deals with the text, the latter with the formal elements that can be isolated from the text. A modified definition of formal correspondence is given to include all those isolable elements that occupy the same places in the economies of the two texts, target and source. The extent to which formal correspondents occur in normal, unrestricted translation is examined, and it is concluded that some

are preserved in all translation, no matter how free. Their proportion is smaller when one takes the perspective of an individual analyst with his specific contrastive topic. In some cases, correspondences are impossible or unlikely, in others they are spurious, and in still others they are a by-product of structural shifts that occur in the course of translation.

An English - Serbo-Croatian contrastive model, based on an English corpus and its translation into Serbo-Croatian, is implicit in the foregoing discussion. Its form is roughly the following: a description of English is taken as given and is exemplified in the English text (corpus), and possibly modified; the text is translated into Serbo-Croatian and the formal correspondents of the isolable elements described by English grammar are sought; a grammatical description of these Serbo-Croatian formal correspondents is made; this description is contrasted with the description of the elements of English with which the analysis began; inferences for learning are drawn.

NOTES

(1) I say this notwithstanding the announcement of the German PAKS project described by G. Nickel and K. H. Wagner, ("Contrastive Linguistics and Language Teaching", IRAL, VI, 3, 1968, 233-55). This project will be based on the transformational model of linguistic description, which is certainly a very exciting prospect, but one has serious qualms about its exhaustiveness in the present state of development of the transformational theory. A case could perhaps be made for a corpus-based analysis along transformational lines.

(2) E. A. Levenston, "The 'Translation-Paradigm', A Technique for Contrastive Analysis", IRAL, III, 3, 1965, 121-25.

L. Spalatin, "Contrastive Methods", SRAZ, 23, 1967, 29-48.

(3) Ž. Bujas, "Concordancing as a Method in Contrastive Analysis", SRAZ, 23, 1967, 49-62.

(4) Such a project is actually already under way; cf. R. Filipović, The Choice of the Corpus for the Contrastive Analysis of Serbo-Croatian and English, in the present volume.

(5) J. C. Catford, A Linguistic Theory of Translation, Oxford University Press, London, 1965, p. 17.

(6) J. Darbelnet, Review of "On Translation" and "Aspects of Translation", The Journal of the Canadian Linguistic Association, VI, 1, 1960, p. 75.

(7) J. C. Catford, op. cit., p. 27.

(8) No operational tests of translation equivalence are offered: ideally, the authority of the translator as "a competent bilingual speaker" (W. Marton, "Equivalence and Congruence in Transformational Contrastive Studies", Studia Anglica Posmaniensia, I, 1-2, 1968, p. 54) is accepted. Speaking more practically, the analyst should be allowed to discard those examples from the corpus in which his own bilingual feeling disagrees. Errors in translation are sufficiently common to make this provision necessary.

(9) The principles of selection of topics are discussed in the 'Research Guide for Project Workers" (for the English - Serbo-Croatian Contrastive Study) prepared by W. Nemser and V. Ivir.

(10) "Languages differ essentially in what they must convey and not in what they may convey." (R. Jakobson, "On Linguistic Aspect of Translation", On Translation, ed. R. A. Brower, first published 1959, Oxford University Press, New York, 1966, p. 236.)

Leonardo Spalatin (Zagreb)

APPROACH TO CONTRASTIVE ANALYSIS

Inter-lingual similarities

The basic assumption of contrastive analysis is that while languages are different, there is always a certain degree of similarity between them. If there were no similarity there would be no contrastive analysis, in the same way as there would be no contrastive analysis if there were no differences. This position falls somewhere half way between that of traditional structuralism, which stresses the uniqueness of each language, and that of transformationalism which stresses their fundamental similarities.

The fact that most of what is written or said in one language can be translated into another language indicates that there must be a certain, rather high, degree of similarity between languages. At the same time translation shows that the similarity is always only partial, that it can never become identity, even with cognate languages or with dialects of the same language.

Languages may be said to consist of some isolable elements and of certain arrangements of the isolated elements. Language elements are assigned to various hierarchical ranks of structural units and to levels[1] according to certain criteria, mostly of a paradigmatic, distributional or extralinguistic nature. If this classification of language elements is carried out by a consistent application of a language theory, the results will show greater or less similarity between languages in regard of the isolated language elements and their properties.

However great the difference in the isolated language elements between languages, it is still possible to render a very large portion of meaning conveyed by the elements of one language into another. This possibility seems to imply two things: (1) similarity between languages is not necessarily limited to similarity between elements belonging to corresponding levels in the languages concerned, and (2) similarity between languages is not necessarily limited to similarity between elements belonging to corresponding classes or ranks in the languages concerned.

The semantic field

Language elements can be assigned to the lexical or the grammatical level. They belong to the lexical level when they are regarded purely from the point of view of their collocational range, regardless of the form with which they enter into a collocation. Thus, the collocations of the words "bark" and "dog" are lexical in "a barking dog /a dog barks/ the barking of a dog /a dog that barks/ the bark of a dog". (Certain arrangements of words, showing grammatical collocations, like "father-in-law", are assignated to the lexical level because they collocate lexically as a unit rather than as the individual elements of which they consist.) A word belongs to the grammatical level when its form is decisive for its collocations with other elements. Thus "I" collocates with "love" but not with "loves". The order of collocations itself is a grammatical feature; thus, "I may not" but not "I not may". To put it in a different way, a word contains a semantic element and a grammatical element. The semantic element determines its collocation with other words; the grammatical element determines its collocations with other grammatical elements. "Peter" does not collocate with "go" not because of its semantic properties (it can collocate with "goes") but because its grammatical properties, i.e. a noun-form without a suffix collocates with a verb-form consisting of the base and -s. The form father can collocate grammatically as "father loves /I see my father/ of the father / this father" but not as * "These father", for which collocation a different form of the word father is required. The semantic element contained in the word father allows it to enter into combinations such as "a father loves / a loving father / a fatherly love / a father's love / a fatherly kiss / a father kisses / a father who kisses / an unnatural father" but not * "a fatherly father". The grammatical, lexical and phonological (probably also the graphemic) levels cover the semantic field of a language.

Similarity at different levels, ranks or classes

Languages can be viewed as covering a certain semantic field. The more similar the culture and civilization of the users of various languages, the greater the resemblance there will be between the semantic fields covered by their languages. But however similar are semantic fields covered by various languages, there will always be areas in one language without counterparts of the same rank or of the same level, or with no counterpart at all, in the other. For the American word "drugstore" there is no equivalent word in

Serbo-Croatian (S-C).

The lack of formal correspondence[2] between two languages at the lexical level does not exclude the possibility of rendering the meaning expressed by an element in one language in the other language, lacking an equivalent element, by means of a different rank or level. The meaning expressed by the American word "drugstore" can be expressed in S-C by a noun-phrase like "a store selling drugs, shirts, cigarettes, etc." (Similar equivalences established at different ranks are widely utilized in monolingual dictionaries.)

The lack of formal correspondence at the grammatical level often has as a consequence a full or partial lack of semantic equivalence. The present-perfect verb-phrase of English (E) and the preterit tense have the same equivalent in S-C. Thus, "He has arrived" and "He arrived" will have their equivalent in the S-C sentence "Došao je". In such cases equivalence is only partial, as only some components of the total meaning are expressed in the target language, in this case the component common to both tenses, that is "past". The component "connected with the present" of the E present-perfect verb-phrase has no equivalent in S-C. In certain contexts the equivalents of the two E tenses are two tenses in S-C. The present "živim" in the S-C sentence "Živim ovdje pet godina" is equivalent to the E present-perfect "I have lived" in the sentence "I have lived here for five years". The S-C perfect "živio sam" in "Živio sam ovdje pet godina" is equivalent to the E preterit "lived" in "I lived here for five years".

Occasionally a grammatical element in one language has no equivalent in another. This is often the case with the articles in E when an E sentence is translated into S-C. The E sentence "He lives in a house" and "He lives in the house" will most frequently have in S-C only one equivalent: "Živim u kući", which means that S-C usually does not distinguish between a marked and an unmarked noun.

Even where there is formal correspondence between two languages, some elements present in a structure of one rank in one language may be represented by different means in the other, although they may have units of the same rank with identical structure. Thus, both E and S-C have noun-phrases with identical structure. The E noun-phrase "His large house, which is near the park..." shows the structure "modifier+modifier+head+qualifier". The corresponding S-C noun-phrase "Njegova velika kuća, koja je blizu parka..."

shows the same structure. On the other hand, the noun-phrase of the clause "A little boy plays" has as its S-C equivalent a clause in which equivalence is established at ranks different from those occurring in the E clause. The modifier "a" of the E noun-phrase "a little boy" has a possible S-C equivalent at the clause level. To convey the meaning of "a" the elements figuring in the structure of the equivalent S-C clause are arranged in such a way that the head of the noun-phrase is placed at the end of the clause. The modifier "little" can have a S-C equivalent at the word level in the form of a bound morpheme giving the noun a diminutive meaning (dječak = boy, dječačić = little boy). Although "little boy" and "dječačić" are of the same rank (because they play the same role in clause structure, that of the subject), the modification is of different types. In E it consists of words arranged according to the rules for the arrangement of a noun-phrase element. In S-C the modification is achieved at word rank by means of bound morphemes (dječač-ić) following morphotactic rules of elements at word rank.

Although three levels of language elements can be isolated in all languages, their semantic coverage varies from language to language. Thus the E emphatic construction "it is... who/that", where emphasis is conveyed by grammatical means (i. e. by a certain distribution of language elements), is rendered in S-C, among other possibilities, by phonological means, that is by giving an emphatic stress to the equivalent of the E word that fills the slot.

The fact that two languages distinguish similar word-classes does not necessarily mean that a word-class in the one language will always have as its equivalent the formally corresponding word-class in the other. Both E and S-C have a word-class usually called adjectives, containing the sub-class of possessive adjectives, which function as modifiers in noun-phrases. Thus "my father" in S-C is most frequently "moj otac", where moj belongs to the same adjective sub-class as the E my. But the clause "My father has arrived" is translated into S-C as "Otac mi je došao" (= father to me has come), where mi is a qualifier belonging to the class of personal pronouns. The sentence can be also translated as "Ø Otac je došao" where zero is equivalent to "my" or "our".

Languages differ in the way they utilize elements of various levels to cover their semantic fields. The verbal aspect of S-C (grammatical level) can in some cases be expressed by various lexical items in E (lexical level): the lexical and grammatical meaning of the S-C imperfective verb "raditi" can

be rendered by the E verb "work", while the lexical and grammatical meaning of the perfective S-C verb "uraditi" can be conveyed by the E verb "accomplish", as in "Radio je mnogo ali uradio malo" = "He worked a lot but accomplished little". Similarly, aspect in S-C can be rendered in E by the selection of a verb for the imperfective aspect and of a different verb with its object for the perfective aspect. The object is derived from the E verb which corresponds to the S-C imperfective aspect:

gledati = look - pogledati = take a look
pjevati = sing - zapjevati =start singing

The following are examples of S-C lexical items having E phrasal equivalents:

starac = old man
starica = old woman
starost = old age
starmali = precocious child
staretinarnica = second-hand shop
starkelja = decreipt old man
ostariti = grow old
starati se (o) =take care (of)

S-C is richer in derivational and form classes than E while E is richer than S-C in the membership of word-classes, so that S-C words often carry a heavier grammatical loan than E words. This situation makes rather dubious Fries's contention that the teaching of E as a second language should concentrate on structures and include only a minimal vocabulary[3]. For the speaker of S-C who tries to learn E this is a rather unsatisfactory method because what is a syntactical feature in E will often be a morphological feature in his language, which difference would interfere with his acquisition of E structure. Often a speaker of S-C will insist on an E word-form only because it is a form with a suffix and thus comes closer to his heavily suffixed native words, preferring "beautifuller" to "more beautiful" or "going" to "go". E compositions written by the students of E at the University of Zagreb, Yugoslavia, show that, at a conservative estimate, lexical mistakes stand to grammatical mistakes in the proportion of 6 to 1.

This high percentage of lexical mistakes reflects the basic difference between the two languages: the heavier reliance of E on lexis which is highly specialized and extremely well developed with items often showing narrow colloca-

tional ranges and utilizing a very large number of specialized bases; and the predominance of grammatical elements in S-C attached to a much smaller body of bases with consequent wider collocations of lexical elements. This can be illustrated by the following example:

dob-ar	dob-r-ot-a	o-dob-ri-ti	dob-ri-čin-a	dob-ra-no
good	good-ness	approve	genial person	consider-ably
dob-r-ostiv	dob-ro-stoje-ći	dob-ro-ćud-an	dob-ro-došao	
grac-ious	well off	mild-temper-ed	wel-come adj	
dob-ro-došlic-a	dob-ro-bit			
wel-come n	well-fare			

As against one base in S-C there are 9 bases in E, and against 26 suffixes and prefixes in S-C there are 8 suffixes in E. (The segmentation may not be precise but still reflects a basic difference between the two languages.)

The high specialization of E lexical items often leads to a situation where for two or more E words there is only one word in S-C; for instance: "light" and "easy" ("lagan"); "difficult" and "heavy" ("težak"); "song" and "poem" ("pjesma"); "turn off" (light), "put out" (fire), "quench" (thirst) ("ugasiti").

The formal correspondence approach to contrastive analysis

Most contrastive analyses employ a formal correspondence approach. We have tried to show that such an approach cannot yield satisfactory results because (1) formal items at one level in one language are not necessarily always equivalent to corresponding formal items in another language, (2) the structure of formal correspondents at a certain rank can be considerably different and the correspondents can still play the same grammatical role in the contrasted languages and convey equivalent meanings, (3) formal correspondence can be simply nonexistent. We shall now try to show, through examples from a very superficial contrastive analysis of some features of E and S-C, what happens when a formal correspondence method is applied, and how misleading and incomplete the results are.

If we accept the formal correspondence approach to contrastive analysis we are confronted with a number of problems. One is how to establish formal correspondences. In most cases they are established on the basis of in-

tuition. We feel that certain formal items play similar roles in the structure of the languages concerned. Thus the is felt to correspond formally to der, die, das in German.

Formal correspondence of the items to be selected for contrastive analysis is often established also on the basis of similar labels. E personal pronouns are contrasted with S-C personal pronouns because of the similar terms used in the respective grammars to designate the two sets of items.

There is a serious question, however, as to whether formal correspondence can offer an effective approach to contrastive analysis. On the basis of what has been said on the foregoing pages, we believe that formal correspondence is far from satisfactory for purposes of contrastive analysis, since it often establishes similarities which are of little practical value, while ignoring subtler forms of similarity which, although they may be less frequent than the formal correspondences, must be taken into account in contrastive research. An example is reflexivity in E and S-C. Both languages have verbs with reflexive objects, and such verbs are felt to be formally correspondent. But statistics show that the S-C reflexive se corresponds in E more often to zero or nothing than to one of the -self forms. Similarity of distribution assigns to the -self forms the role of formal correspondents, but if contrastive analysis stops at this point, the practical result will be that we will have to warn the speaker of S-C against forming his sentences on the evidence of formal correspondence if we do not want him to produce E sentences of the type "I am walking myself" or "I am laughing myself", which do not occur in E. On the other hand, we cannot say that zero or nothing is the formal correspondent, because the S-C sentence "Umiva se" with the reflexive se has as its E equivalent "He is washing himself", in addition to "He is washing". Since instances of se as verb object have as their only phonetically realized E formal correspondents the -self forms, we take the -self forms as the formal correspondents of the S-C se. As for the phonetic zero in E where se occurs in S-C, we are not sure whether what is involved is a zero morpheme (a transitive verb with zero object: "I shave every morning"), or nothing at all. Thus we are faced with formal correspondence with very low equivalence probability. It is obvious that such a formal correspondence will have little practical value resulting from contrastive analysis.

Even in the cases where formal correspondence exists for a large number of instances, there are areas of similarity between the contrasted languages

which are not covered by formal correspondence. Thus my in my father in E corresponds to moj in moj otac in S-C. A similar situation obtains in thousands of other instances. The conclusion is that the forms my, your, etc., have as their S-C formal correspondents the forms moj, tvoj, etc. However, closer scrutiny reveals that while this is true for over fifty percent of cases of my, etc., there still remains a rather high percentage of instances where my, etc., corresponds in S-C to the enclitic dative of personal pronouns, to the reflexive se, to zero, etc. If, for purposes of contrastive analysis, we take into consideration only those items which are formally correspondent in the two languages, no useful contrastive analysis is likely to result because such an analysis will ignore, as often occurs, other similarity relationships which are not formal correspondences in the languages analyzed, although they are equivalences.

The S-C learner of E will react in the same intuitive way and select E possessive adjectives as formal correspondents of S-C possessive adjectives and will, equating formal correspondence with equivalence, produce an impermissibly high percentage of sentences in the target language on the model offered by formal correspondence which will be wrong, the mistakes occurring in the areas where equivalence is no longer valid. In a sentence like "Take (the) hand out of (the) pocket", modeled on S-C "Izvadi Ø ruku iz Ø džepa", with zeros at the place of E possessive adjectives, the formal correspondence "moj, etc. equals my, etc." is no longer operative, and Ø is not intuitively felt to have the function of E possessive adjectives. This means that even in cases where formal correspondence can be established intuitively, or in some other way, the learner will have to be told not to rely on it entirely.

In some cases it is almost impossible to establish any kind of formal correspondence. This is the case with the E article for a speaker of S-C. He is quite helpless before it since there is nothing directly discernible in his language which could serve as a formal correspondent he could utilize when going from S-C to E. In such cases, for all practical purposes we have to give up the formal correspondence approach to contrastive analysis, since the only thing we can contrast is the absence of a set of morphemes in one language with their presence and characteristic distribution in the other. Although no correspondence can be established, it is possible to establish certain equivalence relations on the basis of word-order, demonstrative adjectives and possibly some other elements. A similar situation obtains in respect of S-C verb

aspect and its E equivalences. In such cases even the formal correspondence approach to contrastive analysis has to give up contrasting similar formal categories, and try to establish similarity by means of other value, saying, for instance, that the S-C imperfective "brinuti se" corresponds to E "worry" and the S-C perfective verb "pobrinuti se" to E "see (to something)", which is definitely not a contrastive analysis of the category of aspect in the two languages.

Use of the formal correspondence approach is often due to a misconception of methods of contrastive analysis, which are often confused with methods of language description. This confusion results in the implicit conclusion that if languages are describable in terms of certain categories, contrastive analysis should be in terms of the same categories. This is a falacious assumption because there need not necessarily be any similarity between descriptive methods and contrastive methods. The two are quite independent processes with different aims in view: one discovers and classifies language elements, the other contrasts meanings conveyed by language elements isolated in various languages. It is true that a poor description will yield poor results in contrastive analysis, but not necessarily a poor contrastive analysis. Good methods of description, on the other hand, applied to contrastive analysis are no guarantee of good contrastive analysis.

The semantic approach

Our experience is that languages can be effectively contrasted only on a semantic basis, specifically, on the basis of translation equivalence. The translation approach produces, in addition to one or more high translation probability equivalents, a series of low translation probability equivalents, and the two together cover an entire particular semantic field. Thus, the translation of a corpus containing E possessive adjectives into S-C yields not only possessive adjectives (the result which we get if we accept the formal correspondence approach) but also personal pronouns in their enclitic dative forms, the enclitic form of sebi (i.e. si), the reflexive pronoun (se), possessive adjectives derived from nouns (noun: otac, derived possessive adjective: očev = father's), words like vlastiti, rodjeni (one's own), etc., that is, the whole field of "possessivity"; similarly, an E corpus containing this and that will produce, in addition to ovaj, taj, onaj, a whole series of words containing

the ov-, ta- and on- morphemes (ovdje = here, tamo = there, onamo = yonder, ovako = this way, tako = that way, etc.), the non-productive demonstrative morpheme -s (večera-s = this evening, jesena-s = this autumn, jutro-s = this morning, etc.), that is, the whole field of "demonstrativity".

As with formal correspondence, translation equivalence will not be of great help to contrastive analysis where equivalence is practically nonexistent. But one of the advantages of the semantic approach is that absence of translation equivalence is less frequent than absence of formal correspondence.

Conclusion

Our conclusion is that the formal correspondence approach excludes in advance the possibility of semantic similarities between language elements in various languages which are not beforehand established as formally correspondent. Thus, if we decide that nouns in E and S-C are formally correspondent, and we restrict contrastive analysis to nouns in the two languages, we exclude all instances of other word classes and other linguistic levels which are the only semantic equivalents of certain E nouns in certain environments. Formal correspondence allows for some differences in the distribution of items selected to be contrasted, but it does not allow for similarities among items belonging to different classes, ranks or levels.

We have tried to show that languages differ widely as regards the choice and distribution of the elements they utilize to cover their semantic fields, but that it is comparatively easy to establish semantic equivalences between the fields of various languages, which seems to indicate that contrastive relationships between languages are more profitably established if their equivalent semantic features are compared than if the comparison is based on their equivalent formal elements.

NOTES

1. By "level" we mean one of the main aspects of language, a phonology, grammar or lexicon. The term "rank" refers to hierarchical arrangements of language units. It is usually convenient to isolate five ranks: morpheme, word, phrase, clause and sentence.

2. By a "formal correspondent" we mean a formal element or string of formal elements showing an organized structure at a certain hierarchical rank in one language whose role in the overall language organization is similar to that of some formal element in another language. Thus in both English (E) and S-C, elements can be organized into phrases, functioning as subject or complement at the clause level, consisting of no modifier or one or more modifiers, a head, and no qualifier or one or more qualifiers. Such phrases are formally correspondent in the two languages. Formally correspondent also, for example, are adjectives because in both languages they can function as modifiers of the head of a noun-phrase.

3. "A person has 'learned' a foreign language when he has... first, within a limited vocabulary, mastered the sound system ... and has, second, made the structural devices ... matters of automatic habit". Charles C. Fries, Teaching and Learning English as a Foreign Language, p. 3.

Rudolf Filipović (Zagreb)

THE CHOICE OF THE CORPUS FOR A CONTRASTIVE ANALYSIS OF SERBO-CROATIAN AND ENGLISH

The first problem facing the researchers engaged in the Serbo-Croatian - English Contrastive Analysis Project[1] was that of the method[2], and then immediately after that, closely connected with the selection of the method, came the question of the corpus. Obviously, the choice of the method determined whether a specific corpus was needed or not.

The problem of the method was discussed in the earliest stages of our work on the Project.[3] Professor Pavle Ivić noted that one of the three approaches was possible: traditional, structuralist, or generative.[4] The first approach was discarded as outmoded and unsuitable for contrastive analysis. But Professor Ivić was aware of the difficulties in choosing between the structural and the generative approaches. Our progressive orientation - says Professor Ivić - would dictate the choice of the most modern, i. e. generative-transformational, approach. However, the situation in the field forces us to compromise, "to combine classical structuralism with the elements of the generative approach".[5]

In an article on Serbo-Croatian - English contrastive analysis,[6] I have examined several contrastive studies now available and found that none employs a specific and consistent method that might be regarded as the method of contrastive analysis. Authors of these studies have used different approaches and have come out with results of different quality. My conclusion after reading these studies, supported by our own limited experience, is that there is a strong interdependence between theory and practice in contrastive work, and that the best approach will be a combination of theoretical and empirical methods of research. Our experience so far shows that there are areas of contrastive analysis in which no theoretical method would be successful or would lead to a satisfactory solution. These considerations have prompted us to seek a method, or a combination of methods, that will yield not only theoretical but also practical results. The practical results must be usable in developing better teaching materials and techniques.[7] This will be possible only if they are presented in a form that an average reader of our projected monograph can understand.

In order to fulfil all the requirements of the contrastive description of Serbo-Croatian and English, it was decided to adopt the translation method[8] based on a corpus of examples. This choice then led naturally to another problem: that of the corpus.

Already at the first meeting of the Project (Zagreb, April 1967), I discussed, among other questions, the problem of the corpus.[9] It was decided at that meeting that the corpus should be built upon the following principles: (1) the English corpus should include extracts from the works of both British and American authors; (2) fiction and technical literature should be represented; (3) the corpus should be translated by translators from all the Yugoslav centres represented on the Project, i.e. both the eastern and the western variant of Serbo-Croatian should be equally represented; (4) the original Serbo-Croatian corpus should be built from works of Croatian and Serbian fiction and technical literature and should be translated into English by British and American native speakers, so that both variants of English are equally represented.

Such a corpus would be sufficiently representative on both the English and the Serbo-Croatian side. The translations of the two parts of the corpus (into English and Serbo-Croatian respecitvely) would supply a sufficient quantity of contrastive elements on the basis of which statistically significant conclusions could be made. Two kinds of extracting were envisaged: for a general sample and for a supplementary sample. The general sample would consist of 100,000 sentences and their translations; this was regarded as a sufficiently large corpus for the majority of contrastive elements. (Estimates made on the basis of several frequency studies showed that this corpus would be sufficient for most aspects of syntactic analysis.) The supplementary corpus would be used in those cases in which the general corpus failed to yield a reliable distribution.

However, discussion at the second meeting of the Project[10] indicated that it might not be possible to obtain a sufficiently representative corpus on these principles. Another reason which prompted us to seek a new solution was the fact that it was rather difficult, if not impossible, to build such a large corpus within the limited time and with the resources that we had at our disposal. A whole series of questions would have had to be answered: how to select the works for extracting, how to determine the ratio between British and American (and Croatian and Serbian) texts, between fiction and technical literature, etc.

It became quite clear then that we would have to use an existing corpus and at the same time have recourse to a computer and other data processing devices. There are, at present, two large corpora : one built on British material, spoken and written, and one on American material - only written. We felt that the first corpus, containing material from both spoken and written sources, would better suit our needs and give more significant results.

The first corpus, A Survey of English Usage, has been compiled under the leadership of Professor Randolph Quirk of the Department of English, University College London.[11] A number of studies have been made on the basis of this corpus in recent years: some of them have already been published, some are due to appear soon, and some are in preparation. It is a pity that this corpus has not yet been published as such and is thus practically inaccessible to researchers. Attempts have been made to publish it, but the necessary funds could not be found. For the time being, it is only accessible to a limited number of people, which also limits its usefulness for linguistic research.[12]

The Survey has been compiled on the following three principles: (1) The linguistic material includes all grammatical elements found in the chosen strings of actually recorded written and spoken English. The Survey aims at explaining, in the light of the statistical norm, every grammatical phenomenon regardless of its frequency. Another aim is to describe the conditions under which the speaker deviates from the norm. (2) The Survey embraces the full range of educated English usage, from learned and technical writing to spontaneous colloquial speech. Particular attention is paid to natural (unscripted) speech in different situations - from highly technical discussion to free conversation among friends. Different registers are compared among themselves and with the so-called "literary speech", i.e. dialogue in written literature. (3) The Survey covers present-day English, from 1950 onwards.

In addition to the three main principles, the compilers of the Survey have followed certain other principles: each unit of text consists of about 5,000 words, and the whole corpus includes over 200 texts of British English; the sources include the spoken language, novels, plays, poetry, criticism, and other literary prose, then psychology and social sciences, philosophy, physics and physical sciences, biology, law, politics, religion, then newspaper language, etc. The entire corpus contains over a million words and

is designed to supply sufficient data on high frequency phenomena and most productive norms in the language. Clearly, low frequency structures and certain high frequency variants will not be sufficiently represented in the primary material and supplementary samples will be needed.

By its composition and size, this corpus would meet the requirements of our Project. However, two main reasons have prevented us from taking the corpus of Professor Quirk's Survey of English Usage: (1) The corpus is not readily accessible since it does not exist in printed form; (2) it is not designed for computer processing. Those were the two main reasons behind our decision to work with the American Brown corpus, whose most serious drawback compared with Professor Quirk's corpus is that it does not cover the spoken language.

The Brown Corpus (short for the Standard Corpus of Present-Day English[13]) includes 1, o14, 294 words in connected text of edited English prose published in the United States in the course of 1961. All authors are - as far as it could be established - native speakers of American English. The material included in the corpus was published in the course of a single year, 1961, although parts of it may have been written somewhat earlier.

The corpus is composed of 5oo samples, each containing about 2, ooo words. Each sample begins at the beginning of the sentence, though not necessarily at the beginning of the paragraph in the text from which it is taken; it ends with the sentence in which the two thousandth word appears. That is why some samples have a little more and others a little less than 2, ooo words. The average sample length is 2, o28. 6 words.[14]

The corpus covers a wide range of styles and types of American prose. Poetry is excluded bacause it raises special linguistic problems different from those raised by prose. Plays are also excluded because they do not represent real writing but rather an imagined representation of speech. Fiction is included, but no sample containing over 5o per cent of dialogue has been admitted. The samples have been chosen for their representative qualities and not for certain subjectively determined values.

The word 'standard' in the name of the corpus does not mean a corpus of 'Standard English': it simply indicates that the corpus can be used for comparative studies requiring the same set of data. It can also mean that the corpus is a convenient basis for the compilation and presentation of other sets of data in English and other languages.[15]

The selection of the fields to be covered and the number of samples in different categories have necessarily depended on certain subjective and objective conditions. To a certain extent they depended on the material that could be found in the Brown University and Athenaeum (Providence) Libraries and in the New York Public Library. A list of the main categories to be included was drawn at a meeting at Brown University in February 1963. In addition to the two chief editors, Professors W. N. Francis and H. Kučera, the meeting was also attended by Professor R. Quirk, the editor of the London corpus.

The entire corpus is divided in two main groups:
1. informative prose (374 samples),
2. imaginative prose (126 samples).
The first group has nine categories. The first three of them cover the press: (a) news stories (political, sports, social, late news, financial, cultural) from dailies (33 samples) and weeklies (11 samples)- total 44 samples; (b) editorials and readers' letters in dailies (19 samples) and weeklies (8 samples; (c) reviews (books, theatre, music, ballet) from daily (14) and weekly (3) newspapers - total 17 samples. Seventeen samples come from the field of religion: from books (7), journals (6), and studies (4). Another category contains 36 samples from the field of skills and hobbies: two samples from books and 34 from journals and magazines. The category of popular writings includes 48 samples (23 from books and 25 from magazines). Serious writing, biographies, memoirs and similar texts include 75 samples (38 from books and 37 from periodicals). Thirty samples are included in the category various writings: 24 from government documets, 2 from foundation reports, 2 from industrial reports, and one each from college catalogues and industry organs. The next category, most numerous after the press, is that of scientific and technical literature (80 samples): 12 from natural sciences, 5 from medicine, 14 from social sciences, 15 from political and legal science and education, 18 from the humanities, 12 from technology and engineering.

It would be difficult to contest the number of samples in different categories: the numbers were fixed after a discussion in which each participant made his own proposals. Taken as a whole, the first section, that of informative prose, covers all areas of human activity as they are reflected in written texts. And the ratios between the number of samples in different categories could hardly be different. The heavy emphasis given to the press (98 samples) is a reflection of the importance attached to the realizations of language in

daily and weekly newspapers. The second most numerous field is that of science, with 8o samples and with all fields of science and engineering represented. Since scientific and technical language is perhaps less varied than the language of other forms of informative writing, the number of samples included in the corpus seems quite sufficient. With other categories of writing it is not easy to find objective criteria that would determine the number of examples. It could be said in conclusion that the wide range of fields from which samples have been taken is a reasonable guarantee that no important human activity of which written record exists has been omitted.

Comparing the number of samples in the first section (374) with that of the second (126), one notes that the ratio is 3 : 1, which would mean that the language of informative prose is thought to be three times as important for linguistic analysis as that of literary prose. When one remembers that the main source of material for linguistic investigations in the past used to be belles letres, one cannot fail to appreciate the significance of this corpus and the novel approach that it introduces into the linguistic study od the English language.

Even the composition of the second section, imaginative prose, is novel. The number of samples in different categories shows that no extra weight is given to fiction in the traditional sense of the word (i. e. novels and short stories). Besides, the section includes three categories that have seldom been used as sources of linguistic documentation. The division of imaginative prose into types is rather original. The first category is that of general fiction, represented by 29 samples (2o from novels and 9 from short stories). The same number of samples is found in the next two categories: adventures and westerns (15 samples from novels and 14 from short stories). Then follows love prose (15 samples from novels and 14 from short stories). Twenty-four samples come from detective novels (2o) and short stories (4); science fiction is represented by 6 samples (3 from novels and 3 from short stories) and humour by 9 samples (3 from novels and 6 from essays and similar forms).

This standard corpus of English in now available on magnetic tape for computer processing.[16] The text can be obtained by running the tape through the computer, which will print the entire text in its orthographic form. The text will then have to be translated into Serbo-Croatian, since contrastive analysis is done on the original text and its translation. When the two ver-

sions are thus obtained, contrastive concordances will be produced by the computer, so that each English sentence is matched by its Serbo-Croatian translation. For technical and financial reasons, it has been decided to shorten the Brown corpus by half, translate only this half, and prepare contrastive concordances for it, matching every element of the English text with its Serbo-Croatian translation equivalent.

It was clear to us from the beginning that the contrastive analysis of two languages (e. g. Serbo-Croatian and English) would require two corpora of equal size and composition which would be translated into the respective languages. This would enable us to examine each phenomenon in both languages from the point of view of its translation. This idea had to be given up, however, for several reasons, and it was eventually decided to work with only one corpus and its translation. Obviously, the Brown Corpus and its Serbo-Croatian translation cannot provide all the elements needed in the analysis: we therefore plan to have another, control corpus which will be smaller, and naturally less representative and "standard", than the shortened version of the Brown Corpus. It will consist of a few Serbo--Croatian novels and their translations into English by native speakers of English.

It may be interesting to describe how we propose to shorten the Brown Corpus. One way would be to cut each sample in half, thus hoping that the representativeness of the corpus will not be affected. However, it seems to us that a 2,000 word sample is the shortest possible sample if it is to remain representative of the text from which it is taken. Another possible way would be to reduce the number of samples in categories and sub-categories by half, hoping again that the main features of the corpus will remain intact. This would mean, for instance, getting only 5 samples from daily newspapers and 2 from weeklies in the category "press", sub-category "political news story", instead of the present 10 and 4 samples respectively.

In our shortening we do not follow either of these principles. It has already been said that the London Corpus, which includes the spoken language and plays, would be more suitable for our purpose than the Brown Corpus, which does not contain materials from these sources. But the Brown Corpus does contain dialogues in narrative texts (when dialogues do not cover more than 50 per cent of the total text) which are very useful for the kind of work that

we are doing. An important objective of our Project is to prepare new teaching materials, and these must include not only narratives but also dialogues. This is the reason why the shortening of the Brown Corpus is not done in a linear fashion but rather in such a way that the samples containing dialogues are preserved in each case at the expense of other samples which are less suitable for the contrastive analysis of English and Serbo-Croatian as we envisage it. (Some of the samples would be rather difficult to translate because of their cultural background: this is the case with certain sports texts.) The category of readers' letters to the editor will be kept as it stands because it comes perhaps closest to the free style of expression that we need most in our corpus.

The corpus thus abbreviated will then be translated into Serbo-Croatian. It has already been said that both variants of Serbo-Croatian, eastern and western, will be represented in the translation and that translators will be engaged in all those centres from which Project members come. One third of the corpus will be translated in Belgrade, one third in Sarajevo, and one third in Zagreb. It is thus hoped that the Brown Corpus translation into Serbo-Croatian will display the greatest possible number of features of all variants.

If it is found that our corpus, now consisting of the shortened Brown Corpus and its regionally representative translation plus a smaller control corpus of Serbo-Croatian texts and their English translations, fails to cover all aspects of Serbo-Croatian, a separate analysis will be made of the aspects that have remained uncovered. This analysis will be made by a group of Serbo-Croatian scholars at Novi Sad led by Professor Pavle Ivić.

On the basis of this corpus, processed by IBM 360 computer in the Zagreb Municipal Computer Centre, we shall obtain contrastive concordances of every single element of the English language and its Serbo-Croatian translation. Equivalence will be the main criterion for the selection of structures for contrasting or comparison. Paired texts and their components will be regarded as equivalent when they have been translated from one language into the other.

NOTES

(1) R. Filipović: Jugoslavenski projekt za kontrastivnu analizu srpskohrvatskog i engleskog jezika. Organizacija i zadaci projekta. (The Yugoslav Contrastive Analysis Project - Serbo-Croatian and English. The Organization and Objectives of the Project) Institut za lingvistiku Filozofskog fakulteta Sveučilišta u Zagrebu, Zagreb, 1968, 16 pp.

(2) R. Filipović: "Zašto 'kontrastivna' analiza?" (Why 'contrastive' analysis?) Živi jezici, Belgrade, X, 1-4, 1968, pp. 1-5.

R. Filipović: "Uloga kontrastivne analize u lingvističkom istraživanju" (The role of contrastive analysis in linguistic research), Filološki pregled, Belgrade, VI, 1-2, 1968, pp. 1-11.

(3) The first seminar for Project members was held in Zagreb, April 1 - 3, 1967. Eight papers were read, most of them dealing with the methods of English - Serbo-Croatian contrastive analysis.

(4) P. Ivić: "Osnovni problemi metoda u našem radu" (The main methodological problems in our work). The paper will be published in Project publications and will be entitled "Nekoliko reči o problemima metoda" (Some methodological remarks).

(5) Ib.

(6) R. Filipović: "Contrastive Analysis of Serbo-Croatian and English", SRAZ, Zagreb, 23, 1967, pp. 5-27

(7) R. Filipović: "Pedagoška primjena kontrastivne analize" (Pedagogic applications of contrastive analysis), Pedagoški rad, Zagreb, XXIV, 3-4, 1969, pp. 138-145

(8) I do not wish to discuss the different forms of this method, broad or narrow. Cp. items (6) and (7).

(9) R. Filipović: "Contrastive Analysis of English and Serbo-Croatian" (a detailed description of the Project and plan of work).

(1o) The meeting was held at Novi Sad, November 17 and 18, 1967.

(11) R. Quirk: "Towards a Description of English Usage", Transactions of the Philological Society, Blackwell, Oxford, 1961, pp. 4o-61.

R. Quirk: "On English Usage", Journal of the Royal Society of Arts, 114, London, 1966, pp. 837-51.

(12) Several studies have already been prepared on the basis of this corpus:

(a) R. Quirk: "Descriptive Statement and Serial Relationship", Language, 41, 1965.

(b) R. Quirk: "Types of Deviance in English Sentences", in A Common Purpose, ed. J. R. Squire, Champaign, Ill., 1966.

(c) A. P. Duckworth: "Co-existing Negative Preterite Forms of Dare", Language and Society, Copenhagen, 1961.

(d) A. P. Duckworth, J. Svartvik, and others: "Studies in the Correspondence of Prosodic to Grammatical Features in English", Proceedings of the Ninth International Congress of Linguists, Mouton, The Hague, 1964.

(e) J. Svartvik: On Voice in the English Verb, Mouton, The Hague, 1966.

(f) D. Crystal & R. Quirk: Systems of Prosodic and Paralinguistic Features in English, Mouton, The Hague, 1964.

(g) R. Quirk & J. Mulholland: "Complex Prepositions and Related Sequences", English Studies, 44, 1964.

(h) R. Quirk & D. Crystal: "On Scales of Contrast in Connected English Speech", In Memory of J. R. Firth, Longmans, London, 1966.

(i) R. Quirk & J. Svartvik: Investigating Linguistic Acceptability, Mouton, The Hague, 1966.

(j) H. T. Carvell & J. Svartvik: Computational Experiments in Grammatical Classification, Mouton, The Hague.

(k) D. Crystal & D. Davy: Investigating English Style, Longmans, London.

(l) R. Quirk: "Acceptability in Language", Proceedings of the Philosophical Society, Univ. of Newcastle.

(13) The main information on the sources for this corpus can be found in W. N. Francis, Manual of Information to Accompany A Standard Sample of Present-Day Edited American English for Use with Digital Computers. Department of Linguistics, Brown University, Providence, Rhode Island, 1964.

(14) Wherever the number is over 2,000 words, the surplus material is retained. Because of some computational mistakes, 15 samples have between 1,990 and 1,999 words, and 3 samples have under 1,990 words.

(15) One of the reasons for describing this corpus in such detail is that it might help Yugoslav linguists in preparing a standard corpus of Serbo-Croatian for future linguistic investigations. No standard corpus of this kind has so far been used in writing Serbo-Croatian grammars.

(16) Some studies have already been made on this corpus. Cf. H. Kučera & W. N. Francis, Computational Analysis of Present-Day American English. Brown University Press, Providence, Rhode Island, 1967.

(17) It should be remembered that Professor Quirk's London Corpus contains samples of at least 5,000 words.

Tisak: Kućna tiskara Sveučilišta u Zagrebu - Trg Maršala Tita 14